Investigation of the psychometric properties of the

Inclusion Body Myositis Functional Rating Scale using Rasch analysis

Running head/short title: Rasch analysis of the IBMFRS

Gita Ramdharry PhD [1], Jasper Morrow MBChB PhD [1], Stacie Hudgens MSc [2], Iwona Skorupinska BSc [1], Kelly Gwathmey MD [3], Melissa Currence BA [4], Laura Herbelin BSc [4], Omar Jawdat MD [4], Mamatha Pasnoor MD [4], April McVey MD [4], Richard J Barohn MD [4], Ted M Burns MD [3], Mazen M Dimachkie MD [4], Anthony A Amato MD [5], Michael G Hanna BMBCh MD [1], Pedro M Machado MD PhD [1]

1 Queen Square MRC Centre for Neuromuscular Diseases, Institute of Neurology, University College London, London, UK
2 Clinical Outcomes Solutions, Tucson, AZ, USA
3 Department of Neurology, University of Virginia, VA, USA
4 Neuromuscular research Division, University of Kansas, KS, USA
5 Neurology Department, Brigham and Women Hospital, MA, USA

Acknowledgments

We wish to thank Dr Mike Horton, Director of the Psychometric Laboratory for Health Sciences, University of Leeds, UK for his help and support with the analysis.

Funding acknowledgment, financial disclosures and conflict of interest

The Queen Square MRC Centre for Neuromuscular Diseases is supported by a Medical Research Council grant (MR/K000608/1) [MGH, PMM, GR, JM, IS]. PMM is supported by the National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) at University College London Hospitals (UCLH) NHS Foundation Trust

and University College London (UCL). The views expressed are those of the authors and not necessarily those of the (UK) National Health Service (NHS), the NIHR or the (UK) Department of Health.

This work was supported by a Clinical Translational Science Award (CTSA) grant from National Centre for Advancing Translational Sciences (NCATS) awarded to the University of Kansas for Frontiers: University of Kansas Clinical and Translational Science Institute (# UL1TR002366) [RJB, MMD, MC, LH, OJ, PM, AM]. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH or NCATS.

All authors have no other funding disclosures or conflicts of interest to declare.

Ethical Publication Statement

We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

Abstract word count: 150

Manuscript word count: 3432

Corresponding author:

Dr Gita Ramdharry,

Queen Square MRC Centre for Neuromuscular Diseases

8-11 Queen Square

London

WC1N 3BG

United Kingdom

Abstract

Introduction:

The IBMRFS is a 10 item clinician rated ordinal scale developed for people with inclusion body myositis.

Methods:

Single observations of the IBMFRS were collected from 132 patients. Following Rasch analysis, modifications were made to the scale to optimise fit to the Rasch model, while maintaining clinical validity and utility.

Results:

The original IBMFRS did not fit the assumptions of the Rasch model due to multidimensionality of the scale. Items demonstrated local dependence, disordered step thresholds, and differential item functioning. Deconstructing the scale into upper limb (IBMFRS-UL) and lower limb (IBMFRS-LL) scales improved fit to the Rasch model. A 9 item scale, with the swallowing item removed, (IBMFRS-9) remained multidimensional but demonstrated the ability to discriminate patients along the severity continuum. IBMFRS-UL, IBMFRS-LL and IBMFRS-9 scores were transformed to a 0-100 scale for comparability.

Discussion:

This analysis has led to the development of three optimised versions of the IBMFRS.

Keywords: Inclusion Body Myositis; outcome measurement; Rasch analysis

Introduction

Sporadic inclusion body myositis (IBM) is an acquired muscle disorder associated with ageing, for which there is currently no effective treatment. The Inclusion Body Myositis Functional Rating Scale (IBMFRS) is a quickly administered ordinal rating scale used to determine patients' assessment of their capabilities and independence, developed through modification of the Amyotrophic Lateral Sclerosis Functional Rating Scale (ALSFRS) so creating a disease specific functional scale for IBM by the Muscle Study Group investigators (1). It was more sensitive than other potential IBM outcome measures over 6-month follow up in 28 patients (1). A criticism of the scale is that it was developed by clinicians, rather than from data detailing the patient experience, e.g. interviews. The sIBM Physical Functioning Assessment is a new scale developed from patient interviews and questionnaires (2). Reliability and validity has been established but more extensive psychometric evaluation has not yet taken place to examine the spread and performance of items in a large cohort of patients (3).

The IBMFRS has been extensively used in a number of cross-sectional, natural history and drug trials (4–12). It is also the primary endpoint of the phase 2/3 study of arimoclomol in IBM that started recently (NCT02753530). There is still value in optimising the data obtained from the scale to improve the analysis of current and future trials, and allow improved comparison with previous intervention and non-intervention studies.

Validity has been demonstrated using classic test theory through correlation with other scores, functional assessments and myometry (1,4) but neither inter-rater nor intra-rater reliability has been assessed. Furthermore, unlike the ALSFRS, the IBMFRS has not been adapted/validated to allow self-administration or telephone administration. To be used as an outcome measure, an important consideration is that the IBMFRS is an ordinal scale, where the true sizes of the items

are unknown. These ordinal scores do not support calculations of parametric effect sizes and change scores (13), limiting analysis to non-parametric inferential testing.

Rasch measurement theory is a form of Modern Test Theory (MTT) that postulates the probability of an individual's score on an item of a scale will depend on the difference between the ability of a person and the difficulty of the item, represented on the continuum of the scale. It examines a scale for characteristics and fit to a Rasch model, specifically:

- Uni-dimensionality – whether all items in the scale are measuring the same underlying characteristic. Once the main Rasch factor is taken into account there should be no other associations between items (14).
- Response dependency – whether scores of individual items correlate beyond that expected, helping to identifying redundant items in the scale
- Appropriate category ordering – whether the different responses in each domain are ordered correctly and distinguishable by examiners
- Differential item functioning - whether bias exists for an item among subgroups in the sample, for example gender

The information can be used to interpret the severity hierarchy of individual items as well as item fit and discrimination. Fit statistics test the probabilistic relationship between the expected item performance and item difficulty (13). Additionally, an attribute of Rasch analysis is the development of a linear scale based on the fit of the items to the model using a log odds transformation. This effectively transforms the measurement tool into an interval scale, so more powerful parametric statistical tests can be used to psychometrically or statistically analyse outcomes (15) (14). When used for exploratory rather than confirmatory interpretation, the Rasch model will be supportive of more classical psychometric techniques such as validity, reliability and responsiveness. An overview of Rasch analysis is provided by da Rocha et al. (13) .

We used the Rasch model to examine and understand the individual item and domain level IBMFRS properties as well as ensure the clinical importance of the elements measured were considered at every stage. Our aims were: 1) to assess the overall quality of the IBMFRS using Rasch analysis, and 2) based on this, to develop an optimised IBMFRS using the current items for use in clinical studies.

Methods

IBMFRS

The IBMFRS is a physician administered scale assigning a function score from 0-4 across 10 domains in discussion with the patient for a maximum possible score of 40 reflecting normal function (1) . The 10 domains are relevant for patients with IBM (swallowing, handwriting, cutting food and handling utensils, fine motor tasks, dressing, hygiene, turning in bed and adjusting covers, changing position from sitting to standing, walking, and climbing stairs) and graded on a Likert scale from 0 (being unable to perform) to 4 (normal).

Data source

Data was collected from four sites: University College London, UK; Brigham and Women's Hospital, USA; University of Virginia, USA, University of Kansas Medical Centre, USA. The diagnosis of IBM for each patient was made by a neuromuscular specialist at these institutions, always supported by a muscle biopsy and sometimes other tests such as electromyography and/or muscle imaging.

Procedure:

Multiple observations were recorded from patients with IBM during routine neurology clinic appointments, often at a follow up appointment, and rarely at initial diagnosis. The first, single observation was used for this analysis. The raw scores of the 10 domains were recorded at each site along with the patient age and gender. Age was then categorised for the analysis and was used along with gender and site of data collection to explore if these factors influenced how IBMFRS items performed. The items were scored as described by Jackson et al. (1). Ethical approval for data collection was received prior to data collection from the NHS Research Ethics Committee in the UK, and from the Institutional Review Boards for the US sites. Informed consent was obtained prior to data collection.

istical analysis

Rasch analysis was used to assess the psychometric properties of the IBMFRS scale using Rumm-2030 software (Rumm Laboratory, Australia). Key areas of analysis:

1. *Model fit:* We explored model fit to ascertain the fit of items and persons to the model by reporting the item person interaction fit residual (criteria fit ≤ 2.0). The person's fit residual identifies the degree of divergence between the expected/estimated value and the actual individual person/item data when summed over all persons (per item).
2. *Uni-dimensionality and dependency:* identify any dependency between items and assess whether the scale was unidimensional (r≤0.40)
3. *Category thresholds:* We ascertained whether the thresholds between different responses followed the correct hierarchy of difficulty and were distinguishable as individual responses
4. *Differential item functioning:* We checked for the presence of differential item functioning (DIF) indicating bias between subgroups (significant p-value 0.05)
5. *Individual person response:* If all persons were responding in the way that was expected, we would expect all persons to fall within a fit residual range of ±2.5.
6. *Performance of items:* The item-trait interaction is the main fit statistic to see if the stochastic ordering is the same across the scale.
7. *Person and item separation:* The person separation takes in to account the distribution and targeting of items. We examined the person separation reliability (criteria ≥ 2.0), indicating the internal consistency reliability (criteria ≥ 0.7).

Suggested modifications were made to the scale to meet Rasch model fit expectations. Following this, an interval (logit-based) estimate was established for the Rasch modified scale to allow change scores to be derived from the ordinal categories.

Results

Sample

Data from 132 people with IBM were collected (table 1). The sample was 68% male and the overall mean age was 67.7 (± 8.2) years.

Rasch model

A likelihood-ratio test was undertaken to test the assumption of item homogeneity in the Rasch model and confirm the model to be used in the analysis. The result of the test was highly significant (Chi squared 547.36, $p<0.001$) so a rating scale model was rejected, as equal spacing between thresholds across the trait for each item could not be assumed due to differences in the verbal anchors of the items. A partial credit model for polytomous scales was used as an alternative Rasch model as it allows different thresholds for different items.

Initial analysis of the IBMFRS

The median IBMFRS total score for the sample was 27 (range: 9-39). The data was split into three class intervals by the Rumm2030 software, representing three ability groups. There was reasonable distribution (44:47:41), though fewer than the preferable 50 in each group.

A number of findings with the IBMFRS were identified by the analysis:

1. *Model fit:* The item person interaction fit residual had a high standard deviation of 1.68 around the mean of 0, indicating model misfit. The mean location of person scores was 1.63(±1.53) indicating that participants were of a higher ability than the difficulty level of the scale (item range -2.15 to 2.84). The mean fit residual was -0.25 ±1.06 demonstrating a reasonable fit to the model.
2. *Category ordering:* Three items had disordered thresholds: handwriting, turning in bed and sitting to standing.
3. *Individual person response:* In this sample, the fit residual range was -2.9 to 2.29, with four persons as outliers (less than -2.5).
4. *Performance of items:* The Chi square probability was lower than 0.05 (Chisq 38.55, p<0.008), indicating that some of the scale items may not have been working as expected at grouped levels. Exploration of the individual item fit revealed that the item "swallowing" had a significant Chi square probability, indicating it was not fitting at the adjusted 0.01 level (ChiSq 16.00, p<0.0004). The fit residual for swallowing was highly positive at 3.78 which indicated it was under-discriminating and the item was unrelated to the construct of the overall scale.
5. *Uni-dimensionality:* The IBMFRS appeared to be multidimensional. A principle component analysis of residuals revealed that five items were positively correlated when loaded on the principle component (swallowing, turning in bed, sitting to standing, walking and stair climbing) and five were negatively correlated (handwriting, cutting food, fine motor, dressing, hygiene). When the person ability estimates for those items were compared, a proportion of 0.09 (95% confidence interval [CI] 0.06–0.14) of the t tests performed fell outside the ±1.96 range, indicating multidimensionality. On closer examination, five of the negatively loaded items involved activities of the upper limbs and four of the five positively loaded items involved activities of the lower limbs (with the exception of swallowing).
6. *Item dependence:* Some dependence between items was noted due to high correlations between "walking" and "stair climbing" (r=0.34) plus between "cutting food & utensils" and "fine motor tasks" (r=0.35).

7. *Differential item functioning:* All items worked similarly for all age groups and both sexes. There did, however, appear to be item bias for three items depending on the site the data was collected. The "fine motor tasks" item, "sitting to standing" and "walking" all had significant differences when an ANOVA was performed for the site by class interval interaction.
8. *Person and item separation:* The person separation was 3, with a reliability index of 0.895, meaning the scale could statistically differentiate between 3 groups of patients. The Cronbach's alpha was 0.899 indicating high reliability of the scale items.

<u>Modifications to the IBMFRS:</u>

Changes were made to improve the performance of the scale. Firstly, the item "swallowing" was dropped from the main scale as it did not fit the scale construct, resulting in a 9 item scale (IBMFRS-9). A high proportion of outliers were not present to influence the fit statistics. An alternative justification was the clinical observation of inconsistency sometimes seen between the severity of swallowing difficulties and the general motor impairment (16) (9).

To address the problem of multidimensionality, the scale was split into an upper limb score (IBMFRS-UL) and a lower limb score (IBMFRS-LL) and individual transform scores were calculated for each. From a clinical perspective, upper and lower limb abilities are separate domains functionally and may respond to interventions differently depending on where they are targeted. The items of the upper limb scale were: handwriting; cutting food & utensils; fine motor tasks; dressing and hygiene. The items of the lower limb scale were: turning in bed; sitting to standing, walking and stair climbing.

One item of the IBMFRS-UL and two items of the IBMFRS-LL were rescored to address the disordered thresholds (see table 2). No changes were made to address the differential item functioning by site of data collection, as this is not a fixed demographic and will be highly variable when the scale is in clinical use.

The decision was also made to give the option of a combined 9-item scale (IBMFRS-9), that included all rescored upper and lower limb items, where the effect of an intervention may be more global, and the risks to the model associated with a multi-dimensional scale are deemed to be acceptable.

Performance of the new IBMFRS-UL and IBMFRS-LL :

Both scales demonstrated uni-dimensionality. For the IBMFRS-UL and IBMFRS-LL individually, proportions of 0.015 (CI: 0.022-0.053) and 0.05 (CI: 0.017-0.92) of the t-tests respectively fell outside the ±1.96 range. The IBMFRS-UL showed improved item-person interaction fit residuals, but the standard deviation was still high at 1.45. The IBMFRS-LL was slightly worse with a standard deviation of 1.87. The individual item and individual person fits were improved for both scales, though there was a slight misfit for the item "turning in bed" for the IBMFRS-LL. This may be because the upper limbs can also contribute to this manoeuvre, but the significance was borderline (ChiSq=6.5, p=0.04).

The IBMFRS-UL showed a mean person location much higher than the original scale (3.14 ±2.22), and the participants were of a higher ability than the upper limb scale (item location range: -1.82 to 1.17). The IBMFRS-LL mean person location was much closer to 0 (0.25 ±2.48) indicating that participants were of a similar ability to the lower limb scale (item location range: -1.54 to 2.25). Both scales demonstrated some item gaps, but this was particularly evident for the IBMFRS-LL (figures 1a and 1b). The two lower performing item scores (a score of 0 for "handwriting" and "cutting food & small utensils") on the IBMFRS-UL are redundant with this sample (figure 1b).

The IBMFRS-UL item "handwriting" showed improved ordering of thresholds, as did the two IBMFRS-LL items "turning in bed" and "sit to stand". Figure 2 demonstrates the improvements in threshold ordering.

Splitting the scale resolved the problems of item dependency with no significant correlations between items for the IBMFRS-UL and IBMFRS-UL. In addition, the IBMFRS-UL had no problem with differential item functioning. The item "turning in bed" demonstrated differential item functioning by site, but it was not highly significant (F=3.4, p=0.02).

Performance of the IBMFRS-9:

The 9-item scale retained good overall fit (ChiSq=21.99, p=0.23), but was also multi-dimensional with a proportion of 0.16 (CI: 0.12-0.20) of the t-tests falling outside of the ±1.96 range. Issues of dependency remained, with high correlations between "walking" and "stair climbing" (r=0.32) as well as between "cutting food & utensils" and "fine motor tasks" (r=0.31). The decision was made not to remove any of these items, however, as there is face validity of the items, and they are independent concepts clinically.

The mean persons location was 1.73 (± 1.82), which is slightly higher than the original 10 item scale. The participants remained at a higher ability level than the scale items (item location range -2.88 to 3.05). The items demonstrated a better spread, incorporating the ability of participants, though a small item gap remained at the upper end (figure 1c)

Use of the new scales:

Interval estimates based on the logit scores were produced for the three scales for use in future studies (Supplementary table 1). The transformation table converts the raw ordinal score to interval data, from 0-100, that can be used with parametric statistical analysis.

The choice can be made to use the IBMFRS-UL and IBMFRS-LL scales singly for specific, targeted interventions. For global changes, the IBMFRS-UL and IBMFRS-LL can be used as co-primary endpoints, or the IBMFRS-9 can be used as a single outcome measure.

A modified format for data collection (Table 3) includes options for the new scoring and a prompt to calculate the converted scores from the raw scores. It includes the IBMFRS-UL, IBMFRS-LL and IBMFRS-9 but also an option to score swallowing as a single item if it is deemed important clinically.

Discussion

This pragmatic analysis has led to the development of three optimised versions of the IBMFRS. Interval scales have been developed that can be used to score the scales in trials, with the advantage of meeting the requirements of analysis using parametric inferential tests. A number of steps had to be taken, however, to reach this aim.

If the objective of a scale is to provide a sum score, it is important that all items measure one dimension. The Rasch model is a unidimensional measurement model so the multidimensional nature of the original scale needed to be dealt with by splitting the scale. This leads to improved specificity in the understanding of the presentation of the individual being rated, a clinically useful feature, and the IBMFR-UL and IBMFRS-LL interval scales will allow the use of more powerful parametric statistical analysis. It will, however, require the two scales to be analysed separately when used in trials. If functional performance is the primary endpoint, then the two scales will need to be co-primary outcomes in an IBM study population with arm and leg manifestations. This is a conservative approach to manage the issue of multi-dimensionality, but also gives the option of using one or other scale if an intervention study targets either the upper limbs or lower limbs, e.g. exercise training.

A compromise position would be to use the IBMFRS-9 to measure more global effects of an intervention, though the performance of the scale would be sub-optimal as uni-dimensionality cannot be assumed. The advantage of this optimisation is that as a linear scale there can be better use of IBMFRS data collected clinically, or for research, where comparisons need to be made. Other commonly used clinical tools have been modified in this way, for example the MRC Manual Muscle Testing scale and the modified Fatigue Severity Scale (17,18).

Discarding the “swallowing” item improved the performance of all three scales. It has been noted that the presence of swallowing impairment does not clearly relate to the general level of motor functioning or disease severity, with some speculation that it may represent a different syndrome. Clinical colleagues may still want to rate swallowing as part of an

overall description of disease and functional capacity, but it will not be included and analysed as a combined measure of outcome. It could be analysed as a separate single item tool.

The site data was collected showed difference in how one item functioned on the IBMFRS-LL. Some of the wording of original scale is ambiguous and there is not an extensive manual of how to use the IBMFRS. It would be difficult to justify altering the scale to account for site because this is not a fixed demographic feature, such as age or gender, where different items can be presented according to the demographic category of interest. The site of collection will generally vary. It does, however, highlight the need for standardisation of instructions and alteration of wording to improve clarity.

One of the key weaknesses of all three scales, but particularly the IBMFRS-LL, is the presence of item gaps. Where there are insufficient items to score across the range in abilities, there is a potential reduction in sensitivity to change. The validity of the scale with regard to the patient experience has not been explored, and the next stage of this work is to develop a patient reported scale that can completed face to face, or administered over the phone. This would need to be developed in close collaboration with people with IBM to develop the items and then tested on a sample of patients. This will also be an opportunity to address some of the item gaps identified in this analysis.

It would also be important to understand the contributors to physical, and this could be explored by studying the relationship with other parameters such as quality of life, muscle strength, muscle imaging and other biomarkers.

In conclusion, we have presented a Rasch modified version of the clinician rated IBMFRS, consisting of an upper limb scale, a lower limb scale, and a 9 item combined scale that can be used with the existing item descriptors. An interval

score for each scale has been calculated that can be used in the analysis of the modified IBMFRS when used as an outcome measure.

Abbreviations:

ALSFRS Amyotrophic Lateral Sclerosis Functional Rating Scale

IBM Inclusion Body Myositis

IBMFRS Inclusion Body Myositis Functional Rating Scale

IBMFRS-UL Inclusion Body Myositis Functional Rating Scale, upper limb scale

IBMFRS-LL Inclusion Body Myositis Functional Rating Scale, lower limb scale

IBMFRS-9 Nine item Inclusion Body Myositis Functional Rating Scale

MMT Modern Test Theory

MRI Magnetic Resonance Imaging

References:

1. Jackson C e., Barohn R j., Gronseth G, Pandya S, Herbelin L. Inclusion body myositis functional rating scale: A reliable and valid measure of disease severity. Muscle Nerve. 2008;37(4):473–476.

2. DeMuro C, Lewis S, Lowes L, Alfano L, Tseng B, Gnanasakthy A. Development of the sporadic inclusion body myositis physical functioning assessment. Muscle Nerve. 2016;54(4):653–7.

3. Williams V, Coles T, Gnanasakthy A, Demuro C, Yarr S, Williams N, et al. Psychometric validation of a patient-reported measure of physical functioning in sporadic inclusion body myositis. Muscle Nerve. 2016;54(4):658–65.

4. Lowes LP, Alfano L, Viollet L, Rosales XQ, Sahenk Z, Kaspar BK, et al. Knee extensor strength exhibits potential to predict function in sporadic inclusion-body myositis. Muscle Nerve. 2012 Feb;45(2):163–8.

5. Cortese A, Machado P, Morrow J, Dewar L, Hiscock A, Miller A, et al. Longitudinal observational study of sporadic inclusion body myositis: Implications for clinical trials. Neuromuscul Disord. 2013 May;23(5):404–12.

6. Sancricca C, Mora M, Ricci E, Tonali PA, Mantegazza R, Mirabella M. Pilot trial of simvastatin in the treatment of sporadic inclusion-body myositis. Neurol Sci Off J Ital Neurol Soc Ital Soc Clin Neurophysiol. 2011 Oct;32(5):841–7.

7. Hogrel J-Y, Allenbach Y, Canal A, Leroux G, Ollivier G, Mariampillai K, et al. Four-year longitudinal study of clinical and functional endpoints in sporadic inclusion body myositis: Implications for therapeutic trials. Neuromuscul Disord. 2014 Jul;24(7):604–10.

8. Morrow JM, Sinclair CDJ, Fischmann A, Machado PM, Reilly MM, Yousry TA, et al. MRI biomarker assessment of neuromuscular disease progression: a prospective observational cohort study. Lancet Neurol. 2015 Nov 5;

9. Allenbach Y, Benveniste O, Decostre V, Canal A, Eymard B, Herson S, et al. Quadriceps strength is a sensitive marker of disease progression in sporadic inclusion body myositis. Neuromuscul Disord NMD. 2012 Nov;22(11):980–6.

10. Olthoff A, Carstens P-O, Zhang S, von Fintel E, Friede T, Lotz J, et al. Evaluation of dysphagia by novel real-time MRI. Neurology. 2016 Nov 15;87(20):2132–8.

Ahmed M, Machado PM, Miller A, Spicer C, Herbelin L, He J, et al. Targeting protein homeostasis in sporadic inclusion body myositis. Sci Transl Med. 2016 Mar 23;8(331):331ra41.

Benveniste O, Hogrel J, Annoussamy N, Bachasson D, Rigolet A, Servais L, et al. Rapamycin Vs. Placebo for the Treatment of Inclusion Body Myositis: Improvement of the 6 Min Walking Distance, a Functional Scale, the FVC and Muscle Quantitative MRI [Internet]. ACR Meeting Abstracts. [cited 2018 Nov 5]. Available from: https://acrabstracts.org/abstract/rapamycin-vs-placebo-for-the-treatment-of-inclusion-body-myositis-improvement-of-the-6-min-walking-distance-a-functional-scale-the-fvc-and-muscle-quantitative-mri/

13. da Rocha NS, Chachamovich E, de Almeida Fleck MP, Tennant A. An introduction to Rasch analysis for Psychiatric practice and research. J Psychiatr Res. 2013 Feb;47(2):141–8.

Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum. 2007 Dec 15;57(8):1358–62.

15. Hobart JC, Cano SJ, Zajicek JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. Lancet Neurol. 2007 Dec;6(12):1094–105.

Cox FM, Verschuuren JJ, Verbist BM, Niks EH, Wintzen AR, Badrising UA. Detecting dysphagia in inclusion body myositis. J Neurol. 2009 Dec;256(12):2009–13.

17. van Nes SI, Vanhoutte EK, Faber CG, Garssen M, van Doorn PA, Merkies ISJ. Improving fatigue assessment in immune-mediated neuropathies: the modified Rasch-built fatigue severity scale. J Peripher Nerv Syst JPNS. 2009 Dec;14(4):268–78.

18. Vanhoutte EK, Faber CG, van Nes SI, Jacobs BC, van Doorn PA, van Koningsveld R, et al. Modifying the Medical Research Council grading system through Rasch analyses. Brain J Neurol. 2012 May;135(Pt 5):1639–49.

Figure Legends:

Figure 1: A) Person-item distribution for the IBMFRS-UL. The sample performed better than the scale items with potentially redundant items at the lower end. **B)** Person-item distribution for the IBMFRS-LL. **C)** Person-item distribution for the IBMFRS-9. All three scales have item gaps, as indicated by the red bars, but this is more of a feature for the IBMFRS-LL.

Figure 2: Category probability curves comparing the disordered threshold of the IBMFRS and the improved threshold ordering of the IBMFRS-UL and IBMFRS-LL.

Table 1: Summary of demographics of sample and site of collection.

SITE	Number of participants	Age (years)	Median (IQR) age category 1:40-49, 2:50-59, 3:60-69, 4:70-79, 5:80-89	Sex
University College London	61	66.4 (±8.8)	3 (1)	44 male, 17 female
Brigham and Women's Hospital	26	70.0 (±8.1)	3 (1)	17 male, 7 female
University of Kansas	24	69.6 (±8.2)	3 (1)	16 male, 10 female
University of Virginia	21	66.7 (±5.5)	4 (1)	13 male, 8 female
TOTAL	**132**	**67.8 (±8.12)**	**3 (1)**	**90 male, 42 female**

IQR: interquartile range

Table 2: Item rescoring

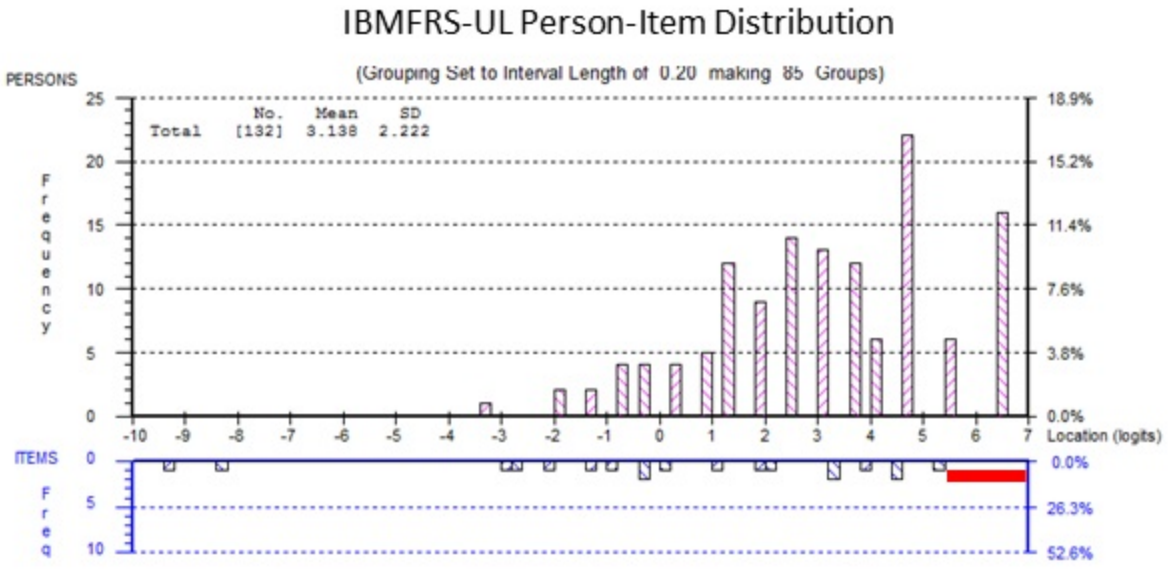
IBMFRS-UL		IBMFRS-LL			
handwriting		turning in bed		sit to stand	
Old score	**New score**	**Old score**	**New score**	**Old score**	**New score**
0	0	0	0	0	0
1	0	1	0	1	1
2	1	2	1	2	2
3	2	3	2	3	2
4	3	4	3	4	3

Table 3: Suggested data collection format for the modified IBMFRS-UL, IBMFRS-LL, IBMFRS-9 and single item swallowing score.

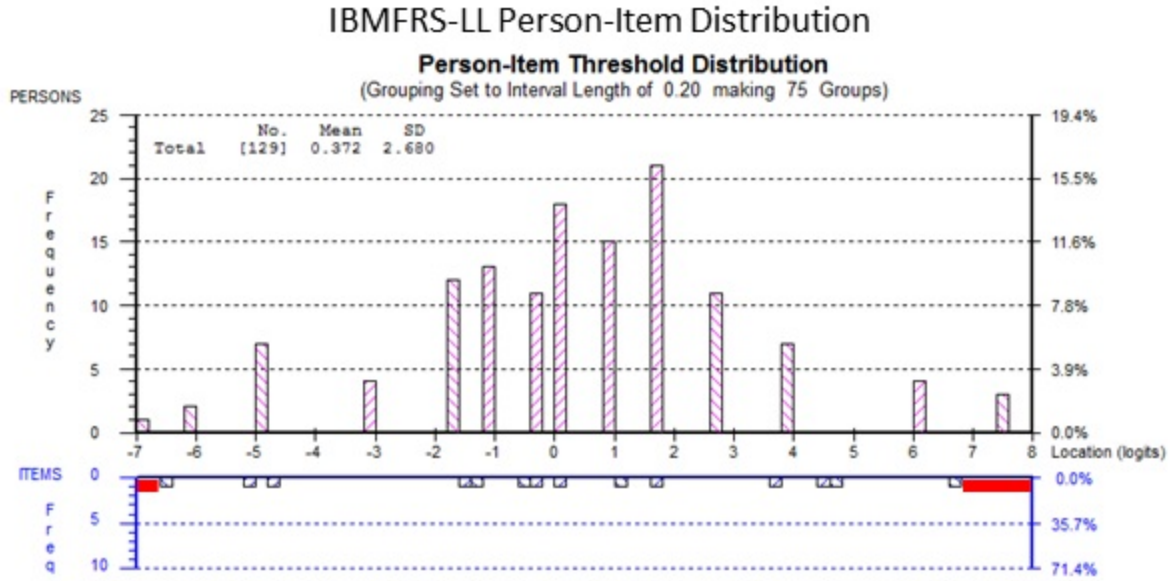
IBMFRS-UL Item	Score and description
1-Handwriting	0- Unable to grip pen *or* able to grip pen but unable to write 1- Not all words legible 2- Slow or sloppy; all words legible 3- Normal
2-Cutting food & small utensils	0- Unable to be fed 1- Food must be cut by someone else, but can still feed self 2- Can cut most foods, but some help or adapted utensils needed 3- Somewhat slow and clumsy; but no help needed 4- Normal
3-Fine motor tasks	0- Unable 1- Frequently requires assistance from caregiver 2- Independent but requires modified techniques or assistive devices 3- Slow or clumsy in completing task (but no assistive devices needed) 4- Normal
4-Dressing	0- Total dependence 1- Requires assistance from caregiver for some clothing 2- independent but requires assistive devices or modified techniques 3- independent but with increased effort or decreased efficiency 4- Normal
5-Hygiene	0- Total dependence

	1- Requires occasional assistance from caregiver 2- Independent but requires assistive devices 3- Independent but with increased effort or decreased efficiency 4- Normal
Total Raw Score **Total converted score**	**/19** **/100**
IBMFRS-LL Item	**Score and description**
1-Turning in bed	0- Unable or requires total assistance *or* can initiate but not turn or adjust sheets alone 1- Can turn alone or adjust sheets, but with great difficulty 2- Somewhat slow or clumsy but no help needed 3- Normal
2-Sitting to standing	0- Unable to stand 1- Requires assistance from a device or person 2- Requires use of arms *or* performs with substitute motions (leaning forward, rocking) but without use of arms 3- Independent (without use of arms)
3-Walking	0- Wheelchair dependent 1- Dependent on an assistive device 2- Intermittent use of an assistive device 3- Slow or mild unsteadiness 4- Normal
4-Climbing	0- Cannot climb stairs 1- Requires assistance from a person or an assistive device additional to the hand rail 2- Dependent on hand rail 3- Slow with hesitation or increased effort; uses hand rail intermittently 4- Normal
Total Raw Score **Total converted score**	**/14** **/100**
Total FRS-9 Raw Score **Total converted score**	**/33** **/100**
Single item scale	**Score and description**
Swallowing	0- Needs feeding tube 1- Frequent choking 2- Dietary consistency changes necessary 3- Early eating problems, occasional choking 4- Normal
Total score	**/4**

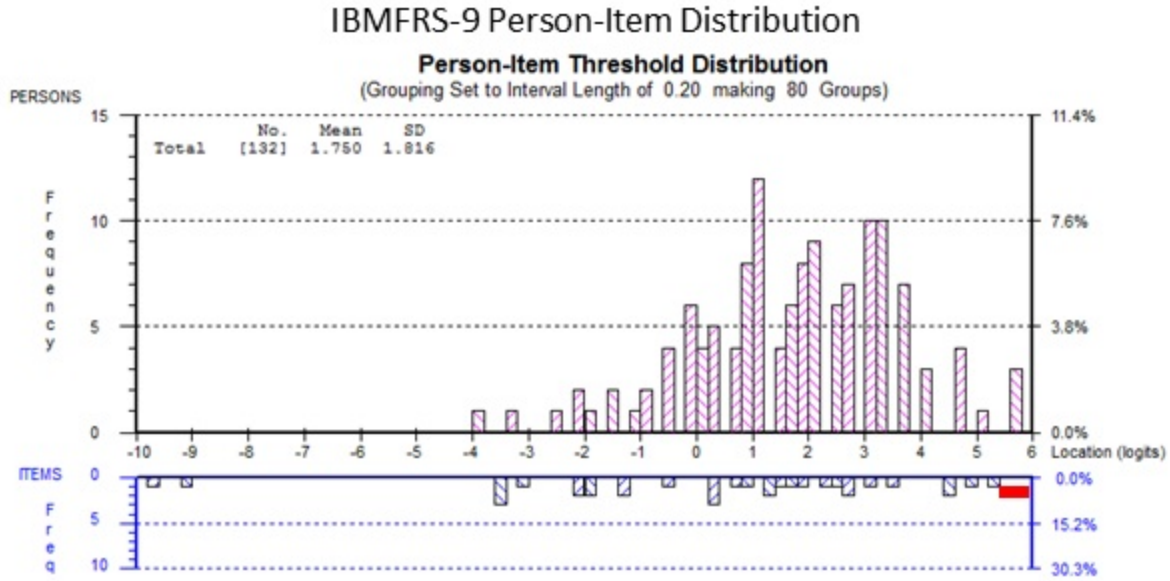
A)



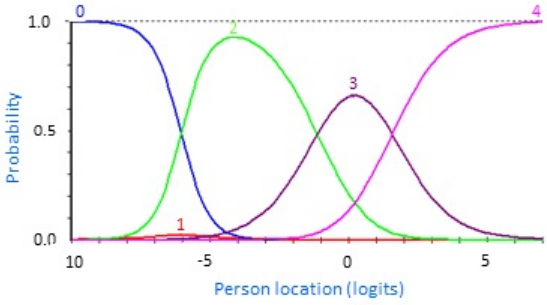
B)



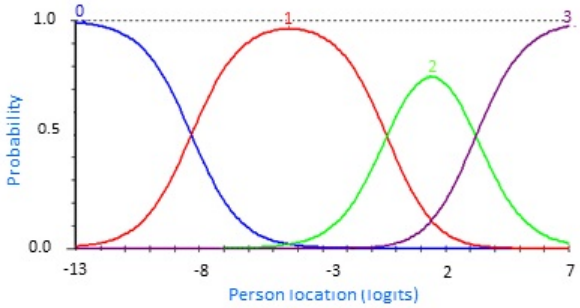
C)



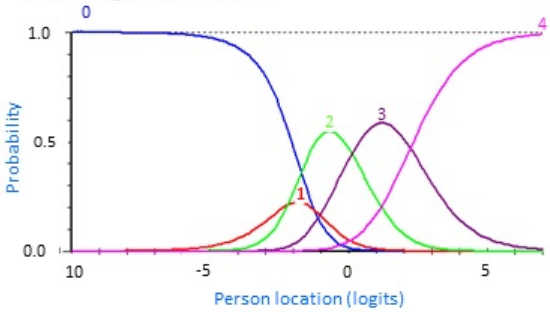
A- Handwriting IBMFRS



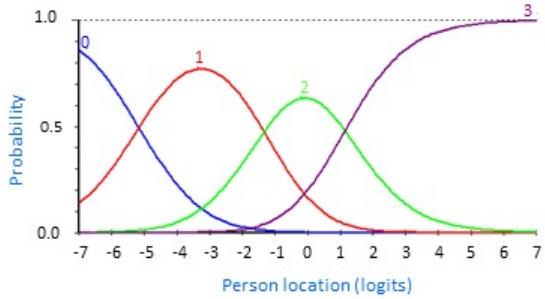
B- Handwriting IBMFRS-UL



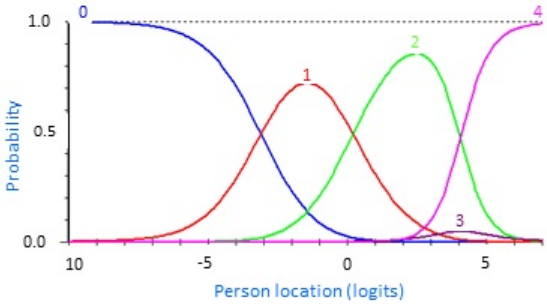
C- Turning in bed IBMFRS



D- Turning in bed IBMFRS-LL



E- Sitting to standing IBMFRS



F- Sitting to standing IBMFRS-LL